

## 知識獲得に向けての品詞体系の考察

出羽 達也<sup>†</sup> 西野 文人<sup>†</sup> 辻井 潤一<sup>‡</sup>

<sup>†</sup>(株)日本電子化辞書研究所 <sup>‡</sup>東京大学

### 1 はじめに

我々は、コーパスからの様々な言語的知識の獲得をコーパスの記述と同時進行的に行なうシステムを開発している。本システムは、獲得された知識や獲得の過程の中間データを格納・管理する中央データベースに対して、各種の言語処理ルーチンがアクセスし、結果を出力するしくみになっている。このようなシステムにおいてシステム全体の整合性を保つためには、中央データベースの入出力インターフェースを統一する必要がある。そのために、入出力の形式をTEI(Text Encoding Initiative)が提唱しているものになるべく準拠したSGML構造で統一する一方(別稿[1]を参照)、記述のための品詞体系の統一を検討している。

望ましい品詞体系は、コーパスから獲得しようとする知識の種類によって異なるが、我々は構文解析結果を利用した知識獲得を計画しているので、構文解析に都合のよい体系でなければならない。本稿では、検討のための題材として、EDR日本語単語辞書[2]の品詞体系(以下、EDRと略す)と、日本語形態素解析システムJUMAN[3]の標準文法の品詞体系(以下、JUMANと略す)の比較を行なう<sup>1</sup>。

### 2 EDR日本語単語辞書とJUMAN標準文法との品詞体系の比較

まず、図1にEDR日本語単語辞書の品詞分類一覧を、図2にJUMAN標準文法の品詞分類一覧を示す。また、図3にはEDR日本語単語辞書の活用の一例としてカ行五段動詞の活用を、図4にはJUMAN標準文法でそれに対応するものとして子音動詞カ行の活用を示す。EDRが基本部分では学校文法に準拠しているのに対し、JUMANは、形容動詞を認めていないこと、判定詞、指示詞といった品詞立て、活用形のし方などに特

<sup>1</sup>類似の試みに奈良先端科学技術大学院大学の羽田らによるもの[4]があるが、これは、語彙拡充の観点から自立語の統合を行なったもので、付属語(接辞を含む)は考慮されていない。

徴がある益岡・田窪文法[5]をベースとしている。

#### 2.1 活用語

活用語の不変化部分に続く部分は、

- (a) 助動詞・助詞
- (b) 派生語尾(接辞)
- (c) 活用語尾

の3通りに考えることができる。この部分に関わるのは、EDRでは、助詞(助詞相当語)、助動詞(助動詞相当語)、接尾語、補助用言、および活用語尾である。一方JUMANでは、助詞、助動詞、接尾辞、および活用語尾である。

JUMANがベースとしている益岡・田窪文法によれば、(a)~(c)は以下のような基準で分類される。

- (a) 一部の例外を除いて、単独で文末の述語の働きができる形式に接続するものに限る。(図4の例で言うと、基本形カタ形に接続するもの)
- (b) 当該の表現が全体として動詞や形容詞に相当する。
- (c) (a)(b)以外。

まず、EDRにおける補助用言はJUMANでは接尾辞に統合されている。さらにEDRで助詞・助動詞と認められていた語の一部が、JUMANでは接尾辞や活用語尾として扱われている。具体的には、EDRで助動詞に分類されているものうち、

「させる」「せる」「られる」「れる」

「がる」「ます」

が動詞性接尾辞に、

「たい」「ない」

が形容詞性接尾辞に、

「た」「う」「よう」

が動詞の活用語尾に分類されている。また、EDRで助詞に分類されているもののうち、

「たら」「たり(だり)」「て」「ば」

が活用語尾として扱われている。そして、助動詞として残ったのは、

「そうだ」「ぬ」「べきだ」「まい」

「みたいだ」「ようだ」「らしい」

などである。

活用のし方については議論の余地があるが、EDRと比べるとJUMANの方が統語的な振舞いに従って品詞が整理される方向にある。例えば、EDRでは他の語の後ろに付いて全体で動詞として振舞う語のうち、「めく」など名詞の後ろにつくものは接尾語に、「すぎる」など動詞の連用形の後ろにつくものは補助用言に、「てしまう」など複数の単語の連続から成って一つの語のようなはたらきをするものは助動詞相当語に分類されていた。これに対してJUMANでは、すべて動詞性接尾辞として扱われる(ただし「てしまう」の場合は、「て」までが活用語尾で「しまう」が動詞性接尾辞)。

しかし、JUMANには(a)の分類基準に見られるように、まだ形態的な観点が残っている。我々の立場では、助動詞として残っているものに対しても統語的な観点を徹底する必要がある。後述するように、本稿では助詞の問題について深く立っていないので、今後助詞とあわせて残りの助動詞の扱いを検討する。

なお、接尾辞の細分の仕方に関しては、「何に付くか」という点は省いて、名詞性接尾辞、形容詞性接尾辞、動詞性接尾辞といった分類の仕方が適当と思われる。

## 2.2 判定詞

判定詞(「だ」「である」「です」)もEDRにはない品詞である。EDRでは助動詞だが、名詞と結合して述語を作ることからJUMANでは独立の品詞としている。統語的な観点からは、JUMANの分類の方が望ましい。

## 2.3 時相名詞/時詞

EDRは時詞という名前、JUMANは時相名詞という名前、副詞としても振舞い得る名詞を分類している。しかしこのような分類は、名詞か副詞のどちらかであるという曖昧性の表現の仕方としては変則的である。我々は、解析結果の曖昧性をそのまま残してコーパスを記述するという方針をとっているため、品詞名に曖

昧性を持ち込むのではなく、他の曖昧性と同じ形で表現したい。すなわち、時相名詞/時詞という品詞分類は廃止し、普通名詞と副詞の両方の見出しを持たせる方がよい。

## 2.4 数量名詞、相対名詞

益岡・田窪は、特殊な性質を持つ名詞として数量名詞と形式名詞をあげている。しかしEDRにも、益岡・田窪文法に基づいているJUMANにも、数量名詞の分類は見られない(時相名詞または時詞に含まれている)。一部の副詞から修飾され得る(例えば、「ずいぶん大勢」という統語的な特徴から、数量名詞の分類は必要と思われる。同様に、「少し左」のように副詞から修飾され得る相対名詞の分類も必要だろう。

## 2.5 形式名詞

JUMANでは、EDRで形式名詞に分類されている語のうち、副詞相当句や副詞節を作る「ため」「よう」などを独立させて副詞的名詞とし、それ以外の「つもり」「はず」などを形式名詞としている。このように分けることは、統語的な観点から望ましいものだが、時相名詞/時詞の項で述べたように「副詞としても振舞い得る名詞」とすべきではない。形式名詞に対して形式副詞とでも言うべきものを立てるべきであるが、形式副詞という名前は奥津[6]が助詞についての議論で用いているもので、これについては助詞の問題とあわせて再度検討する必要がある。

## 2.6 形容詞と形容動詞

EDRでは別品詞としている形容詞と形容動詞を、JUMANでは、活用型の違いとみなして形容詞にまとめている。統語的な振舞いの観点からは品詞を分ける必要性は見当たらないので、この点ではJUMANの方が我々の目的に適っている。

## 2.7 指示詞

指示詞は、JUMANのみにある品詞で、「これ」のような名詞形態、「こう」「こんなに」のような副詞形態、「この」「こんな」のような連体詞形態に分けることができる。これらはEDRではそれぞれ、普通名詞、副詞(または形容動詞連用形)、連体詞(または形容動詞連体形)の一部である。

構文解析の段階では指示詞を独立の品詞とする必要性は特に見当たらないので、この点ではEDRの方が適していると言える。

## 2.8 接頭語／接頭辞

EDR, JUMANともに細分を行なっているが、接頭語／接頭辞は一般に派生語の品詞のあり方に影響しないので、構文解析の段階では細分情報は必要ないと思われる。

## 2.9 その他

副詞は数、種類共に多いので、JUMANが行なっているような細分類はあまり現実的でないように思われる。一方、前述したように数量名詞と相対名詞を設けるならば、それに対応して数量名詞修飾副詞と相対名詞修飾副詞が必要になるだろう。

助詞については、今回は残念ながら十分な検討に至らなかったが、「構文論の中で品詞論に然るべき位置を与えるのがよい」[6]とする奥津による助詞の解体・再編成の議論が参考になると考えている。

また、接続詞については、EDRでは文接続詞と単語接続詞という分類を行なっているが、益岡・田窪文法の定義では「接続詞は、文頭において、先行する文とのつながりを示す役割を果たす。」とあり、EDRが単語接続詞としているものの記述がない。JUMANはこの定義に従って、EDRの単語接続詞に相当する語を助詞に分類している。前述の助詞の議論とあわせて、今後詳細に検討したい。

## 3 まとめ

本稿では、EDR日本語単語辞書と日本語形態素解析システムJUMANの標準文法の品詞体系を比較しながら、構文解析に適した品詞体系について考察した。

今後は、今回十分に踏込めなかった助詞についての検討を行ない、網羅的かつ整合のとれた体系にしていく必要がある。また、EDR日本語単語辞書の全見出し語について、新たに策定する体系との対応表を作成・公開する予定である。

## 参考文献

[1] 西野 文人, 杉山 健司, 辻井 潤一: 「知識ベース増殖のための中央データベース」言語処理学会第3回年次大会予稿集,1997

[2] 日本電子化辞書研究所: 「EDR 電子化辞書使用説明書」1995

[3] 松本裕治, 黒橋禎夫, 山地 治, 妙木 裕, 長尾 真: 「日本語形態素解析システムJUMAN version 3.1 使用説明書」京大工学部長尾研究室, 奈良先端科学技術大学院大学松本研究室,1996

[4] 羽田ゆかり, 松本裕治: 「複数辞書の統合的利用のための汎用日本語辞書の構築」情報処理学会研究報告NL116-1,pp.1-6,1996

[5] 益岡隆志, 田窪行則: 「基礎日本語文法—改定版—」くろしお出版,1992

[6] 奥津敬一郎, 沼田善子, 杉本武: 「いわゆる日本語助詞の研究」凡人社,1986

図1. EDR日本語単語辞書の品詞分類

| [類]  | [品詞名]   |
|------|---|
| 名詞   | 普通名詞<br>固有名詞<br>数詞<br>時詞<br>形式名詞                  |
| 動詞   | 動詞  |
| 形容詞  | 形容詞   |
| 形容動詞 | 形容動詞  |
| 副詞   | 普通副詞<br>陳述副詞                                      |
| 連体詞  | 連体詞   |
| 接続詞  | 文接続詞<br>単語接続詞                                     |
| 接頭語  | 形容詞的接頭語<br>副詞的接頭語<br>連体詞的接頭語<br>接頭小辞<br>前置助数詞     |
| 接尾語  | 接尾語<br>単位<br>後置助数詞                                |
| その他  | 助詞<br>助詞相当語<br>助動詞<br>助動詞相当語<br>補助用言<br>感動詞<br>記号 |

図2. JUMAN 標準文法の品詞分類

空白

| [品詞]     | [細分品詞]    |         |
|----------|-----------|---------|
| 名詞       | 普通名詞      |         |
|          | サ変名詞      |         |
|          | 固有名詞      |         |
|          | 地名        |         |
|          | 人名        |         |
|          | 数詞        |         |
|          | 形式名詞      |         |
|          | 副詞の名詞     |         |
| 時相名詞     |           |         |
| 動詞       |           |         |
| 形容詞      |           |         |
| 副詞       | 様態副詞      |         |
|          | 程度副詞      |         |
|          | 量副詞       |         |
|          | 頻度副詞      |         |
|          | 時制相副詞     |         |
|          | 陳述副詞      |         |
|          | 評価副詞      |         |
|          | 発言副詞      |         |
|          | 接続詞       |         |
|          | 連体詞       |         |
|          | 指示詞       | 名詞形態指示詞 |
| 連体詞形態指示詞 |           |         |
| 副詞形態指示詞  |           |         |
| 接頭辞      | 名詞接頭辞     |         |
|          | 動詞接頭辞     |         |
|          | イ形容詞接頭辞   |         |
|          | ナ形容詞接頭辞   |         |
| 接尾辞      | 名詞性述語接尾辞  |         |
|          | 名詞性名詞接尾辞  |         |
|          | 名詞性名詞助数辞  |         |
|          | 名詞性特殊接尾辞  |         |
|          | 形容詞性述語接尾辞 |         |
|          | 形容詞性名詞接尾辞 |         |
| 動詞性接尾辞   |           |         |
| 判定詞      |           |         |
| 助動詞      | 格助詞       |         |
|          | 副助詞       |         |
|          | 引用助詞      |         |
|          | 名詞接続助詞    |         |
|          | 述語接続助詞    |         |
|          | 終助詞       |         |
| 感動詞      | 句点        |         |
|          | 読点        |         |
|          | 括弧始       |         |
|          | 括弧終       |         |
|          | 記号        |         |

図3. EDR 日本語単語辞書の活用 (例)

・カ行五段活用

|    |       |
|----|-------|
| 未然 | か (こ) |
| 連用 | き (い) |
| 終止 | く     |
| 連体 | く     |
| 仮定 | け     |
| 命令 | け     |

図4. JUMAN 標準文法の活用 (例)

・子音動詞カ行

|         |      |
|---------|------|
| 基本形     | く    |
| 未然形     | か    |
| 意志形     | こう   |
| 命令形     | け    |
| 基本条件形   | けば   |
| 基本連用形   | き    |
| タ形      | いた   |
| タ系推量形   | いたろう |
| タ系条件形   | いたら  |
| タ系連用テ形  | いて   |
| タ系連用タリ形 | いたり  |
| タ系連用チャ形 | いちゃ  |
| 音便条件形   | きゃ   |